

Jahid Hasan

[✉ hasan89@purdue.edu](mailto:hasan89@purdue.edu) | [📞 765-694-5992](tel:765-694-5992) | [LinkedIn Jahid Hasan](#) | [GitHub jahid674](#) | [Website](#) | [📍 Lafayette, Indiana](#)

SUMMARY

My research delivers data-centric, reliability-first pipelines that bridge theory and real-world application. I focus on developing practical automation techniques for data preparation and pipeline optimization, ensuring robust machine learning performance. Currently, I'm seeking an internship to apply these approaches in real-world contexts, broaden my perspective through hands-on collaboration, and contribute meaningfully to impactful, production-scale systems.

EDUCATION

Doctor of Philosophy (Ph.D.) — Computer and Information Technology

Purdue University, West Lafayette

Supervisor: [Dr. Romila Pradhan](#)

AUG 2023 - MAY 2027

GPA: 3.97

Master of Science (M.S.) — Mathematics

University of Dhaka, Dhaka, Bangladesh

Dissertation Title: *Age Distributed model of Communicable Disease in Heterogeneous Environment*

Supervisor: [Dr. Md. Shahidul Islam](#)

Co-Supervisor: [Dr. Md. Kamrujaman](#)

JAN 2022 - JUL 2023

GPA: 3.81

Bachelor Science (B.S.) — Mathematics

University of Dhaka, Dhaka, Bangladesh

JAN 2017 - DEC 2021

GPA: 3.69

SKILLS

Programming Languages/Frameworks: **Python, MATLAB, Seaborn, SQL, R, Fortran, SAS, SPSS**

Machine Learning & AI Tools/Libraries: **Scikit-Learn, PyTorch, Tensorflow, Hugging Face Transformers**

Data & Visualization: **Pandas, NumPy, MySQL, MongoDB, Matplotlib, Seaborn, Tableau**

WORK EXPERIENCE

Purdue University, West Lafayette, IN — Graduate Research Assistant

AUG 2023 - PRESENT

- Designing a system that randomized data inputs and sources to amplify diversity and originality in LLM-generated responses—demonstrating how harnessing randomness transforms it from noise into a powerful driver of uniqueness.
- Built a framework—PIPELENS to optimize the pipelines (Database and ML) for unseen data by leveraging data profiles and historical execution. Currently under submission, it demonstrates up to **20%** superior performance over baselines —[LEARN2CLEAN](#), [BUGDOC](#).
- Developed a framework that assesses and explains the robustness of the optimized ML pipeline in drifted circumstances. Moreover, the framework automates suggestions for a new pipeline efficiently—bypassing the need for retraining when the current one fails. [1]
- Architected DATASIFT, a reinforcement-learning framework for training-data selection that outperforms [AUTODATA](#), RANDOM, and ENTROPY baselines over **10 to 35%**, delivering fairer and more accurate models through uncertainty-aware data valuation. [2]
- Leading technical efforts in the [NSF](#) and [CASMI](#)-granted project *Data Preparation for Fair and Trustworthy Machine Learning*.

University of Dhaka, Bangladesh — Graduate Research Assistant

JAN 2022 - JULY 2023

- Conducted rigorous dynamical-systems derivations and executed comprehensive data analysis for the *Age-Distributed Communicable Disease* project supported by the National Science and Technology Department, Bangladesh.

INVOLVEMENT

- **CIT Alternate Senator**, Purdue University (2024–Present). Representing my department in the Graduate Government, voicing student concerns, facilitating communication, and contributing to legislative and policy decisions affecting graduate students.
- **Event Director**, Polytechnic Graduate Student Organization (2024–2025). Co-founded this organization for 1,100+ techie graduate students at Purdue by drafting bylaws and building its structure, while leading events, budgets, and logistics that made it popular among students.
- **Fundraising Chair**, Bangladesh Student Association at Purdue (2024–2025). Led fundraising strategy that secured \$56,000 from Purdue's Student Fee Advisory Board (SFAB) to support cultural exchange activities, while organizing events and strengthening donor relationships to advance the organization's mission.

AWARDS & PROFESSIONAL ACTIVITIES

- Awarded **US Microsoft EDBT Fellowship**, USA, 2025.
- Participant, International Conference on Management of Data 2025, **ACM SIGMOD PODS 2025**, Berlin, Germany, 2025.
- Attendee, EDBT Summer School on AI and Data Management, Nicosia, Cyprus, 2025
- Travel grant for the workshop on **Digital Safety**, The Center for Advancing Safety of Machine Intelligence, UL Research Institute. May 2024.
- National Science and Technology Fellowship (NST), Bangladesh, 2022. A fellowship by the Science and Technology Ministry of Bangladesh.
- Merit Scholarship, University of Dhaka, Bangladesh, 2022. A scholarship for 95th percentile academic results.
- Reviewer in Elsevier 2023 & 2024, External Reviewer in VLDB, SIGMOD, CIKM, EDBT, [ORCID](#)

PUBLICATIONS

[1] Jahid Hasan and Romila Pradhan. Explanations for machine learning pipelines under data drift. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, HILDA '25, New York, NY, USA, 2025. Association for Computing Machinery.

[2] Jahid Hasan and Romila Pradhan. Data acquisition for improving model fairness using reinforcement learning. *arXiv preprint arXiv:2412.03009*, 2024.